FEATURE WEIGHTING IN K-MEANS CLUSTERING

ABSTRACT

A method and system is provided for integrating multiple feature spaces in a k-means clustering algorithm when analyzing data records having multiple, heterogeneous feature spaces. The method assigns different relative weights to these various features spaces. Optimal feature weights are also determined that lead to a clustering that *simultaneously* minimizes the average intra-cluster dispersion and maximizes the average inter-cluster dispersion along *all* the feature spaces. Examples are provided that empirically demonstrate the effectiveness of feature weighting in clustering using two different feature domains.

5

10

ARC9-2000-0078   37